IMPROVING ACCURACY IN FRAUD DETECTION IN E-COMMERCE USING NOVEL NEURAL NETWORKS OVER RANDOM FOREST

R. Pavithra[1], K. Somasundaram[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: To detect the fraud in E-Commerce Platform based on Novel Neural Networks and Random Forest Algorithms.
Materials and Methods: The performance analysis for maximum accuracy in Fraud detection using Neural Network (N=10) over Random Forest Algorithm which identifies fraud in E-Commerce Platform. GPower is used to compute sample size using a pretest power of 0.8 and an alpha of 0.05.
Result: Mean accuracy of Novel Neural Networks 94.54% is high compared to Random Forest 93.21%. Significance value for accuracy and loss is 0.421 (p>0.05).
Conclusion: When compared to Random Forest, the accuracy of Novel Neural Networks is higher.

Keywords: Machine Learning, Novel Neural Networks, Random Forest, Accuracy, Fraud Detection, E-Commerce.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105. [2*]Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India. Pincode: 602105.

1. Introduction

Electronic trade, sometimes known as internet business, is a strategy that allows businesses and individuals to buy and sell items through the internet (Cao et al. 2021). Massive volumes of data have been stored and moved from one place to the next in recent years, due to the Internet and E-commerce (Gilchrist 2017). While migrating data, Fraudsters can make use of information from the customer (Saputra and Suharjito 2019). But Extortion is often recognised after it has occurred in many organizations. Extortion identification is the ideal choice for destroying it from the climate and preventing a repeat if they can't prevent it in a timely manner (Lebichot et al. 2021). The applications of Fraud Detection in E-Commerce are banking, financial services and telecommunications (Daliri 2020).

In the last 5 years, there have been 118 articles in IEEE Explorer and 152 in ResearchGate. Synthetic Minority Over-inspecting Technique (SMOTE) measure is to be utilized to make balance information. Consequences of assessment utilizing disarray lattice accomplish the most noteworthy exactness of the Novel Neural Networks (Luo and Wan 2019; Saputra and Suharjito 2019). The accuracy of the markers in the recognition model, which is enhanced to perceive the misrepresentation exchanges from the authentic ones with an approximate accuracy of 83 %, is confirmed using a real-world dataset (Luo and Wan 2019; Paasch 2008). This work aims to propose an acceptable framework for eliminating the characteristics of extortion trade, such as individual and exchange-related pointers. There are indeed two items included in it: item type and item nature. The two components significantly improve the accuracy of extortion detection (Zheng et al. 2018). The use of behavior-based techniques to detect online payment fraud has been recognized as a viable method. However, using low-quality behavioral data to generate high-resolution behavioral models is difficult (Wang and Zhu 2020). This paper introduces TradaBoost, a new transfer learning algorithm that improves on TradaBoost. It takes a more thorough look at the weights of incorrectly classified cases (Zheng et al. 2020). This work builds a model for the problem of class imbalance that includes a trade-off between sensitivity and accuracy, but also a Big Data-driven ecosystem, and puts it to the test using large-scale data (Makki et al. 2019).Our team has extensive knowledge and research experience that has translated into high quality publications(Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022)

The limitation distinguished from the current framework is low precision. This review is to improve the accuracy of misrepresentation recognition by fusing Machine Learning algorithms like Novel Neural Networks and Random Forest. The aim of this model is to improve and detect misrepresentation in online business stage.

2. Materials and Methods

This review was made in the Software Computing Lab, Department of Information Technology, Saveetha School of Engineering. Test size has been determined utilizing Gpower software by contrasting both the controllers. Two gatherings are chosen for contrasting the measure and their outcome is determined. Test size is 10. This is to alter the issue of low precision rate Novel Neural Networks and Random Forest calculation is utilized. Mean accuracy of Novel Neural Networks is 94.54%. Mean accuracy of Random Forest algorithm is 93.21%. Dataset for this article is collected from the website https://github.com/abdul-random/Fraud_Detection_E-Commerce (abdul-random n.d.) with 151113 columns and 11 rows (Sharma et al. 2019).

The proposed work is planned and executed with the assistance of Python OpenCV software.The stage to evaluate profound learning was Windows 10 OS. Equipment setup was an Intel center i3 processor with a RAM size of 4GB. Framework sort utilized was 64-bit. For execution of code, Python programming language was utilized. The dataset is used behind the scenes during code execution to simulate a yield interaction for precision.

Novel Neural Networks

A brain neuron gets an input and produces an output that is utilized by another neuron based on that input. In learning about acquired data and then predicting outcomes, the neural network simulates this process. The network is made up of three layers of configurations and connections of neurons. The input layer is the first and only layer that receives signals from the outside world. Signals from the input layer are received by the neurons in the following layer, known as the hidden layer. From the incoming signals, the hidden layer extracts significant characteristics or patterns. The key features or patterns are then routed to the output layer, which is the final layer of the network. Pseudocode for Novel Neural Networks is shown in Table 1.

Random Forest

The supervised learning method is used by Random Forest, a well-known machine learning algorithm. It can be used for both classification and regression problems in machine learning. It is based on ensemble learning, which is a way of combining numerous classifiers to solve a complex problem and improve the performance of the model. Random Forest is a classifier that averages the results of a number of trees on different subsets of a dataset to improve the dataset's predicted accuracy. It forecasts the final output based on the majority votes of predictions and the predictions from each tree. The more trees in the forest, the more precise it is, and the problem of overfitting is avoided. Pseudocode for Random Forest is shown in Table 2.

Statistical Analysis

SPSS software is used for statistical analysis of Novel Neural Networks and Random Forest. Independent variables are user id, signup time, purchase time, purchase value, device id, source, browser, sex, age, ip address, class. Dependent variables are lower bound ip address, upper bound ip address, accuracy. Independent T test analysis is carried out to calculate accuracy for both methods.

3. Results

In statistical tools, the total sample size used is 10. This data is used for analysis of Novel Neural Networks and Random Forest algorithm. Statistical data analysis is done for both the prescribed algorithms namely Novel Neural Networks and Random Forest algorithm. For the purpose of identifying fraud in E-Commerce, the group and accuracy values are determined. The 10 data samples utilized for each algorithm, as well as their losses, are used to create statistical values that can be compared. Table 5 shows that group, accuracy and loss values for two algorithms Novel Neural Networks and Random Forest algorithms are denoted. Group statistics table displays the number of samples gathered. Mean and standard deviation obtained and accuracies are calculated and entered.

Table 6, shows group statistics values along with mean, standard deviation and standard error mean for the two algorithms are also mentioned. Independent sample T test is applied for data set fixing confidence interval as 95%. Table 7, shows independent t sample tests for algorithms. The comparative accuracy analysis, mean of loss between two algorithms are specified. Fig. 1, shows comparison of mean of accuracy and mean loss between Novel Neural Networks and Random Forest algorithm.

Mean, standard deviation and standard error mean for Novel Neural Networks are 94.5400, 1.38929, 0.43933 respectively. Similarly for Random Forest, the mean, standard deviation and standard error mean are 93.2180, 1.24497, 0.39369 respectively. On the other hand, the loss values of Neural Network for mean, standard deviation and standard error mean are 5.4600, 1.38929, 0.43933 respectively. For Random Forest, the loss values of Random forest for mean, standard deviation and standard error mean are 5.7820, 1.24497 and 0.39369 respectively. The group statistics value along with mean, standard deviation and standard error mean for the two algorithms are also specified. The graphical representation of comparative analysis, means of loss between two algorithms of Novel Neural Networks and Random Forest are classified. Fig. 1, shows comparison of mean of accuracy and mean loss between Novel Neural Networks and Random Forest algorithm. When compared to the 93.21% of Random Forest, Novel Neural Networks outperforms it by 94.54%.

4. Discussion

In the given study, the significance value obtained is 0.421 (two-tailed, $p>0.05$) which shows that Novel Neural Networks appears to be better than Random Forest. Table 3, shows that accuracy analysis of the Novel Neural Networks classifier is 94.54% whereas Table 4, shows the accuracy of the Random Forest classifier is 93.21%.

The calculation neural network is a man-made consciousness strategy whose idea is to apply a neural organization framework in the human body where hubs are associated with one another (Hollmén, Lagus, and Soininen 2000). Random Forest is a calculation utilized in the order of a lot of information (Shaohui et al. 2021). It is an advancement of the Classification and Regression Tree strategy by applying the bootstrap collecting technique (Hollmén 2000). Random Forest is used to blend of every great exchange misrepresentation tree which is then joined into one model. Random Forest depends on an irregular vector esteem (Blokdyk 2018).

The limitations of this study is that it takes a very long time to train a neural network, especially with large datasets. The future scope of this study is that the system should be expanded to include a larger number of datas with lesser time consumption in training the data set.

5. Conclusion

Based on this study, the mean accuracy of Random Forest is 93.21% whereas Novel Neural Networks have a higher mean accuracy of 94.54%.

Hence it is inferred that Novel Neural Networks appeared to be better in accuracy when compared to Random Forest.

Declarations

Conflicts of Interest

No conflict of interest in this manuscript.

Authors Contribution

Author RP was involved in data collection, data analysis and manuscript writing. Author KS was involved in conceptualization, data validation and critical reviews of manuscripts.

Acknowledgements

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.
1. BestEnlist Company, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

abdul-random. n.d. "GitHub - Abdul-random/Fraud_Detection_E-Commerce: Fraud_Detection_E-Commerce Is a Project in Which We Have to Predict Which Transaction Is Fraud on the Basis of Various Online Customer Attributes." Accessed September 30, 2021. https://github.com/abdul-random/Fraud_Detection_E-commerce.

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102142.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." Bioresource Technology 351 (May): 126956.

Blokdyk, Gerardus. 2018. Mobile Fraud Detection Third Edition. 5starcooks.

Cao, Chenhong, Yi Gao, Yang Luo, Mingyuan Xia, Wei Dong, Chun Chen, and Xue Liu. 2021. "AdSherlock: Efficient and Deployable Click Fraud Detection for Mobile Applications." IEEE Transactions on Mobile Computing. https://doi.org/10.1109/tmc.2020.2966991.

Daliri, Sajjad. 2020. "Using Harmony Search Algorithm in Neural Networks to Improve Fraud Detection in Banking System." Computational Intelligence and Neuroscience 2020 (February): 6503459.

Gilchrist, Alasdair. 2017. Tackling Fraud: Behavioural Biometric Analysis: Fraud Detection in Finance, Business and ECommerce.

Hollmén, Jaakko. 2000. User Profiling and Classification for Fraud Detection in Mobile Communications Networks: Dissertation.

Hollmén, Jaakko, Krista Lagus, and Timo Soininen. 2000. User Profiling and Classification for Fraud Detection in Mobile Communications Networks.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." Environmental Research. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." Materials Chemistry and Physics. https://doi.org/10.1016/j.matchemphys.2022.125900.

Lebichot, Bertrand, Theo Verhelst, Yann-Ael Le Borgne, Liyun He-Guelton, Frederic Oble, and Gianluca Bontempi. 2021. "Transfer Learning Strategies for Credit Card Fraud Detection." IEEE Access. https://doi.org/10.1109/access.2021.3104472.

Luo, Suyuan, and Shaohua Wan. 2019. "Leveraging Product Characteristics for Online Collusive Detection in Big Data Transactions." IEEE Access. https://doi.org/10.1109/access.2019.2891907.

Makki, Sara, Zainab Assaghir, Yehia Taher, Rafiqul Haque, Mohand-Said Hacid, and Hassan Zeineddine. 2019. "An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection." IEEE Access. https://doi.org/10.1109/access.2019.2927266.

Paasch, Carsten A. W. 2008. Credit Card Fraud Detection Using Artificial Neural Networks Tuned by Genetic Algorithms.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." Journal of Energy Storage. https://doi.org/10.1016/j.est.2022.104422.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Saputra, Adi, and Suharjito. 2019. "Fraud Detection Using Machine Learning in E-Commerce." International Journal of Advanced Computer Science and Applications. https://doi.org/10.14569/ijacsa.2019.0100943.

Shaohui, Du, Guanwen Qiu, Huafeng Mai, and Hongjun Yu. 2021. "Customer Transaction Fraud Detection Using Random Forest." 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE). https://doi.org/10.1109/iccece51280.2021.9342259.

Sharma, Vidish, Tarun Trehan, Rahul Chanana, and Suma Dawn. 2019. "StudieMe: College Recommendation System." 2019 3rd International Conference on Recent Developments in Control, Automation & Power Engineering (RDCAPE). https://doi.org/10.1109/rdcape47089.2019.8979030.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." Computational Intelligence and Neuroscience 2022 (February): 5906797.

Thanigaivel, Sundaram, Sundaram Vickram, Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" Fuel. https://doi.org/10.1016/j.fuel.2022.123494.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." International Journal of Biological Macromolecules 209 (Pt A): 951–62.

Wang, Cheng, and Hangyu Zhu. 2020. "Representing Fine-Grained Co-Occurrences for Behavior-Based Fraud Detection in Online Payment Services." IEEE Transactions on Dependable and Secure Computing. https://doi.org/10.1109/tdsc.2020.2991872.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Zheng, Lutao, Guanjun Liu, Chungang Yan, and Changjun Jiang. 2018. "Transaction Fraud Detection Based on Total Order Relation and Behavior Diversity." IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/tcss.2018.2856910.

Zheng, Lutao, Guanjun Liu, Chungang Yan, Changjun Jiang, Mengchu Zhou, and Maozhen Li. 2020. "Improved TrAdaBoost and Its Application to Transaction Fraud Detection." IEEE Transactions on Computational Social Systems. https://doi.org/10.1109/tcss.2020.3017013.

TABLES AND FIGURES

Table 1. Pseudocode for Novel Neural Network

```
//I : Input dataset records
Import required packages.
Convert data sets into numerical values after the extraction feature.
Assign data to X train, Y train, X test and Y test variables.
Using train_test_split()function, pass training and testing variables.
Give test_size and random_state as parameters for splitting data using the Neural Network training model.
Compiling model using matrices as accuracy.
Calculate accuracy of model.
OUTPUT//Accuracy
```

Table 2. Pseudocode for Random Forest

```
//I : Input dataset records
Import required packages.
Convert data sets into numerical values after the extraction feature.
Assign data to X train, Y train, X test and Y test variables.
Using train_test_split()function, pass training and testing variables.
Given test_size and ‘n_estimaors’ : [10, 20, 100], ‘max_depth’ : [2, 4, 6, 8] as parameters for splitting data using the Neural Network training model.
Compiling model using matrices as accuracy.
Calculate accuracy of model.
OUTPUT//Accuracy
```

Table 3. Accuracy of Fraud Detection in E-Commerce using Novel Neural Networks

Test size	**Accuracy**
Test 1	92.29

Test 2	93
Test 3	93.18
Test 4	94.01
Test 5	94.67
Test 6	95
Test 7	95.34
Test 8	95.56
Test 9	95.65
Test 10	96.70

Table 4. Accuracy of Fraud Detection in E-Commerce using Random Forest

Test size	**Accuracy**
Test 1	92
Test 2	92.89
Test 3	93.12
Test 4	93.49
Test 5	94.56
Test 6	94.89
Test 7	95
Test 8	95.17
Test 9	95.50
Test 10	95.56

Table 5. Group, Accuracy, Loss value uses 8 Columns with 8 width data for Fraud Detection in E commerce.

S.NO	**Name**	**Type**	**Width**	**Decimal**	**Columns**	**Measure**	**Role**
1	Group	Numeric	8	2	8	Nominal	Input
2	Accuracy	Numeric	8	2	8	Scale	Input
3	Loss	Numeric	8	2	8	Scale	Input

Table 6. Group Statistical Analysis of Novel Neural Networks and Random Forest. Mean, Standard Deviation and Standard Error Mean are obtained for 10 samples. Novel Neural Networks have higher mean accuracy and lower mean loss when compared to Random Forest.

	Group	**Algorithm**	**N**	**Mean**	**Std. Deviation**	**Std.Error Mean**
Accuracy	1	NEURAL NETWORK	10	94.5400	1.38929	0.43933
	2	RANDOM FOREST	10	93.2180	1.24497	0.39369
Loss	1	NEURAL NETWORK	10	5.4600	1.38929	0.43933
	2	RANDOM FOREST	10	5.7820	1.24497	0.39369

Table 7. Independent Sample T-test: Confidence interval as 95% and level of significance as 0.05. Novel Neural Networks is insignificantly better than Random Forest with p value 0.421 (p>0.05).

		F	**Sig.**	**t**	**df**	**Sig(2-tailed)**	**Mean difference**	**Std.Error Difference**	**Lower**	**Upper**
Accuracy	Equal variances assumed	0.050	0.421	0.54	18	0.032	0.32200	0.58992	-0.91738	1.56138
	Equal variances not assumed	-	-	0.54	17.788	0.032	0.32200	0.58992	-0.91844	1.56244
Loss	Equal variances assumed	0.050	0.421	-0.54	18	0.032	-0.32200	0.58992	-1.56138	0.91738
	Equal variances not assumed	-	-	-0.54	17.788	0.032	-0.32200	0.58992	-1.56244	0.91844

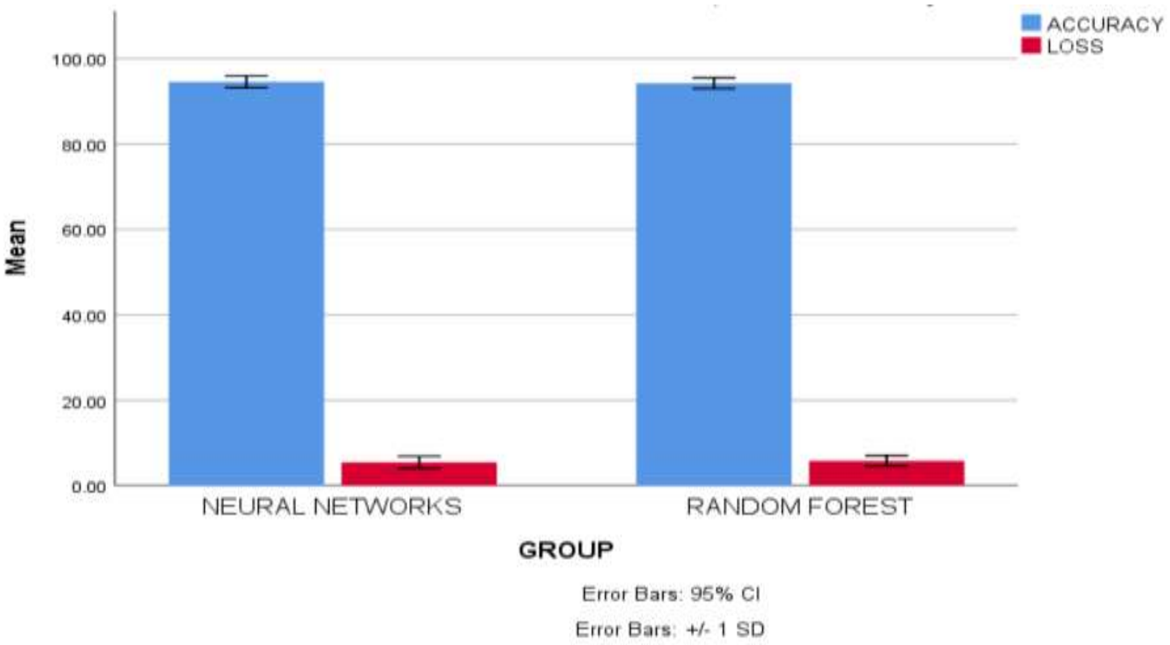

Fig. 1. Comparison of Novel Neural Networks and Random Forest Classifier in terms of mean accuracy and loss. The mean accuracy of Novel Neural Networks is better than Random Forest Classifier. Standard deviation of Novel Neural Networks is slightly better than Random Forest. X Axis: Novel Neural Networks Vs Random Forest Classifier and Y Axis: Mean accuracy of detection ± 1 SD.